

Voice Recognition (VR) B Technology Status

(File VR Brief.wpd, Update 06/28/00)

The ability of computers to recognize and act on human speech is called Voice Recognition (VR). It has been around since the 1980's but has only had limited impact as a computer-human interface. It is technology that seemingly is always just around the corner, but always improving.

Command & Control	Current VR excels in these applications where there are a limited number of voice commands which are appropriate for a specific task.
VR Advantages	VR eliminates the keyboard and potentially is easier to use and faster.
Continuous Speech	Most modern systems do not require the user to speak one word at a time, but can handle continuous normal speech.
Speaker Dependent	These systems are tuned to specific speakers and require brief training sessions where the user speaks pre-defined phrases into the system to tune it. Accuracy is better than speaker dependent systems.
Speaker independent	These systems have the ability to recognize any speaker's voice without training. VR is moving in this direction. Heavily accented speech is still a problem for most speaker independent systems.
System Examples:	<p>Jupiter B from MIT's Spoken Language Systems Group is a speaker independent speech engine to provide weather information via telephone. See: Wwww.sls.les.mit.edu</p> <p>Jupiter recognizes about 2,000 words & runs on a 500 MHz Pentium III.</p>
Example:	Mya , from Motorola, Phoenix, AZ, is designed as a talking Internet voice portal. Tellme Networks, Mountain View, CA has a similar strategy.
Limited Context Domain	Keeps the vocabulary to a manageable level.

Natural Language Processing Known as NLP, this AI technique allows the software to interpret the meaning of the words. Present NLP works best with Limited Context Domains. Outside of this, the technology doesn't do well at present.

Lernout & Hauspie: **Lernout & Hauspie**, Ieper, Belgium, recently purchased Dragon Systems, Inc. is the leading supplier of VR software. Microsoft owns part of L&H.

L&H low-end speech engine B speaker-independent recognition of up to 100 words B requires less than 200 KB of memory & less than 100 Mips of CPU power. Many 32-bit embedded processors (ARM7) can deliver this for less than \$20.

L&H mid-range speech engine B Vocabulary up to 1,000 words. Good for command & control applications.

L&H high-range speech engine B Vocabulary up to 20,000 words. The average adult has a 10,000 word vocabulary. Requires about 5-minutes training for each new user. Can run on reasonable modest processors.

Accuracy (recognition rate) Even the best engines only have a recognition rate of 95-98%

Software learning Today's programs can learn from their mistakes, improving their accuracy over time. More powerful processors & increased memory aid this process since VR is a real-time process.

Increased computer Power This is driving success, although proponents always maintain that VR is not processor intensive. A RISC chip with a few megabytes of memory and a hard drive can provide a serviceable interface for less than \$20. Success, however, seems directly related to processor enhancements.

Information Appliances Prime candidates for voice interfaces. These require command & control VR and are limited context domain applications. Examples: VCRs, microwave ovens, fax machines, cell phones. Any device with more than a few buttons can benefit from VR.

PDA's	PDA's are good candidates for VR because data input is a problem. VR with 2000 word vocabulary could handle all command & control, common words, names and URL needs of a PDA
Automobiles	The AHands off® safety requirement for autos is idea for VR solutions for radio, Internet, navigation, cell phones. VR is currently available for Jaguars and Lincolns. Look for Palm V car dock that features L&H speech engine & 200-word vocabulary.
PC's	May be the last devices to embrace VR. Currently VR software is available for Windows navigation and some dictation. (IBM's Via Voice & Dragon's Naturally Speaking).
Microphones	A long standing problem with VR has been development of noise cancellation microphones which allow VR input without having the mic directly next to the user's mouth. Headset requirements have delayed many potential applications.
Hands-Free Approach	<p>DSP speech processing systems are now being designed to turn cellular handsets into two-way speaker-phones. Targeted for in-car use, this approach uses DSP, echo cancellation, full duplex operation and filtering of background noise. See Analog Devices ADSP-21ESP202 chip which combines a 16-bit, 40 Mips DSP speech processor with voice-codec, hands free speech recognition software. (EET, 6/00)</p> <p>Also see STMicroelectronics (partnered with Lernout & Hauspie), Lucent Technologies (partnered with Philips Speech Processors).</p> <p>Developments here are being driven by State laws which will force hands free cell phone operation while driving.</p>
Technology Advances	<u>IBM's Microdrive</u> , a half dollar size hard drive which offers up to 1 GB of storage, may lead the way toward voice recognition on handheld computers. Current PDA's with 1-4 MB of storage are not adequate for

VR. Several hundred MB of storage is needed for reliable VR. (EET 6/26/00).

Sources: EE Times, 6/5/2000, 6/19/2000